Please amend the claims as follows (this listing of claims replaces all prior versions):

1. (Currently Amended) A <u>computer implemented</u> method of determining whether a set of nucleotides is within a first nucleic acid sequence, the method comprising:

receiving a first and second nucleotide of a second nucleic acid sequence, the second nucleotide being a nucleotide after the first nucleotide;

combining said first and second nucleotide <u>in sequence</u> into a ~~sequential~~ <u>first</u> set <u>of nucleotides</u>; [[and]]

<u>at a computer,</u> comparing the first set of nucleotides to a first nucleic acid sequence to determine whether the first set of ~~sequential~~ nucleotides is within the first nucleic acid sequence<u>;</u> <u>and</u>

<u>if the first set of nucleotides is not within the first nucleic acid sequence, storing the first set of nucleotides as a unit in a database in one or more storage devices for the second nucleic acid sequence.</u>

2. (Canceled)

3. (Original) The method of claim 1, wherein if the first set of nucleotides is within the first nucleic acid sequence, receiving a third nucleotide of the second nucleic acid sequence, the third nucleotide being a nucleotide after the second nucleotide.

4. (Original) The method of claim 3, further comprising: combining the first set of nucleotides with the third nucleotide to make a second sequential set.

5. (Currently Amended) The method of claim 4, further comprising: comparing the second set of nucleotides to [[a]] <u>the</u> first nucleic acid sequence to determine whether the second set of sequential nucleotides is within the first nucleic acid sequence.

6.     (Original) The method of claim 5, wherein if the second set of nucleotides is not within the first nucleic acid sequence, storing said second set as a unit in a database for the second nucleic acid sequence.

7.     (Original) The method of claim 6, further comprising: determining the sum of all units stored for the second nucleic acid sequence.

8.     (Currently Amended) The method of claim 7, further comprising: determining the difference between total number of units stored for [[a]] the first nucleic acid sequence and the total number of units stored for the second nucleic acid sequence.

9.     (Original) The method of claim 8, further comprising: utilizing the difference to determine the distance between the first nucleic acid sequence and the second nucleic acid sequence.

10.     (Currently Amended) A computer readable storage medium comprising instructions that when executed by a machine, causes the machine to perform: the method of claim 1.

identify a first nucleic acid sequence;

receive a first and second nucleotide of a second nucleic acid sequence, the second nucleotide being a nucleotide after the first nucleotide;

combine the first and second nucleotide in sequence into a first set of nucleotides;

compare the first set of nucleotides to the first nucleic acid sequence to determine whether the first set of nucleotides is within the first nucleic acid sequence; and

if the first set of nucleotides is not within the first nucleic acid sequence, store the first set of nucleotides as a unit in a database for the second nucleic acid sequence.

11.     (Currently Amended) A computer implemented method of creating a database of nucleotide units for a first nucleic acid sequence, the method comprising:

receiving a first nucleotide of a first nucleic acid sequence;

at a computer, determining whether the first nucleotide has been stored in a database in

one or more storage devices as a unit for the first nucleic acid sequence; [[and]]

if the first nucleotide has not been stored in the database, storing the first nucleotide as

[[a]] an individual unit for the first nucleic acid sequence;

if the first nucleotide has been stored in the database, receiving a second nucleotide of the

first nucleic acid sequence, the second nucleotide being a nucleotide after the first nucleotide;

combining the first and second nucleotides into a sequential set;

at the computer, determining whether the sequential set has been stored in the database as

a unit for the first nucleic acid sequence; and

if the sequential set has not been stored in the database, storing the sequential set as a unit

in the database for the first nucleic acid sequence.


12 - 14.      (Canceled)


15.      (Currently Amended) The method of claim [[14]] 11, wherein if the sequential set

has been stored, receiving a third nucleotide of the first nucleic acid sequence, the third

nucleotide being the next sequential nucleotide after the second nucleotide.


16.      (Canceled)


17.      (Currently Amended) The method of claim [[16]] 11, further comprising:

determining the sum of all units stored for the first nucleic acid sequence.


18.      (Currently Amended) A computer readable storage medium comprising

instructions that when executed by a machine causes the machine to: the method of claim 11

receive a first nucleotide of a first nucleic acid sequence;

determine whether the first nucleotide has been stored in a database in one or more

storage devices as a unit for the first nucleic acid sequence;

if the first nucleotide has not been stored in the database, store the first nucleotide as an individual unit for the first nucleic acid sequence;

if the first nucleotide has been stored in the database, receive a second nucleotide of the first nucleic acid sequence, the second nucleotide being a nucleotide after the first nucleotide;

combine the first and second nucleotides into a sequential set;

determine whether the sequential set has been stored in the database as a unit for the first nucleic acid sequence; and

if the sequential set has not been stored in the database, store the sequential set as a unit in the database for the first nucleic acid sequence.


19.     (Currently Amended) A system for determining whether a set of nucleotides is within a first nucleic acid sequence, the system comprising:

a data processor executing instructions to implement

a receiving component for receiving a first and a second nucleotide of a second nucleic acid sequence, the second nucleotide being a nucleotide after the first nucleotide;

a combining component for combining said first and second nucleotide in sequence into a ~~sequential~~ first set of nucleotides; [[and]]

a comparing component for comparing the first set of nucleotides to a first nucleic acid sequence to determine whether the first set of sequential nucleotides is within the first nucleic acid sequence;

a storing component for storing said first set as a unit in a database for the second nucleic acid sequence if the first set of nucleotides is not within the first nucleic acid sequence.


20.     (Canceled)


21.     (Currently Amended) The system of claim [[20]] 19, comprising: a second receiving module for receiving a third nucleotide of the second nucleic acid sequence if is determined that the first set of nucleotides is within the first nucleic acid sequence.

22.     (Currently Amended) A method of determining the distance between two nucleic acid sequences ~~sequence~~, the method comprising:

determining the number of words in a first nucleic acid sequence;

combining the first sequence with a second nucleic acid sequence to make a combined nucleic acid sequence;

determining the number of words in the combined nucleic acid sequence; and

determining the difference between the number of words in the combined nucleic acid sequence and the first nucleic acid sequence to determine the distance between the first nucleic acid sequence and the second nucleic acid sequence.

23.     (Currently Amended) A computer readable <u>storage</u> medium comprising <u>instructions that when executed by a machine cause the machine to:</u> ~~the method of claim 22~~

<u>determine the number of words in a first nucleic acid sequence;</u>

<u>combine the first sequence with a second nucleic acid sequence to make a combined nucleic acid sequence;</u>

<u>determine the number of words in the combined nucleic acid sequence; and</u>

<u>determine the difference between the number of words in the combined nucleic acid sequence and the first nucleic acid sequence to determine a distance between the first nucleic acid sequence and the second nucleic acid sequence.</u>

24.     (New) The method of claim 1, comprising generating a first dictionary of words that can be used to build the first nucleic acid sequence, each word comprising at least one nucleotide, and determining a distance between the first and second nucleic acid sequence based on a first number of words in the second nucleic acid sequence that is not in the first dictionary.

25.     (New) The method of claim 24, comprising generating a second dictionary of words that can be used to build the second nucleic acid sequence, each word comprising at least one nucleotide, and determining the distance between the first and second nucleic acid sequence

also based on a second number of words in the first nucleic acid sequence that is not in the second dictionary.

26.     (New) The method of claim 25, comprising determining the distance between the first and second nucleic acid sequence based on a maximum of the first number and the second number.

27.     (New) The method of claim 26, comprising determining a normalized distance based on the maximum of the first number and the second number divided by a maximum of a third number and a fourth number, the third number representing the number of words in the first dictionary, and the fourth number representing the number of words in the second dictionary.

28.     (New) The method of claim 25, comprising determining the distance between the first and second nucleic acid sequence based on a sum of the first number and the second number.

29.     (New) The method of claim 28 comprising determining a normalized distance based on the sum of the first number and the second number divided by a number of words that is needed to build a third nucleic acid sequence comprising the second nucleic acid sequence appended to the first nucleic acid sequence.

30.     (New) The method of claim 28 comprising determining a normalized distance based on the sum of the first number and the second number divided by an average of a third number and a fourth number, the third number representing the number of words that is needed to build a third nucleic acid sequence comprising the second nucleic acid sequence appended to the first nucleic acid sequence, the fourth number representing the number of words that is needed to build a fourth nucleic acid sequence comprising the first nucleic acid sequence appended to the second nucleic acid sequence.

31.     (New) The method of claim 1, comprising determining a distance between the

first and second nucleic acid sequences based on a distance measure for nucleic acid sequences

that satisfies triangle inequality, such that a distance between the first and second nucleic acid

sequences is no greater than a sum of a first distance between the first nucleic acid sequence and

a third nucleic acid sequence, and a second distance between the second and third nucleic acid

sequences.